HAL
open science

# Modèles de Markov cachés crédibilistes en présence d'informations non distinctes

Emmanuel Ramasso, Michèle Rombaut, Denis Pellerin

## To cite this version:

Emmanuel Ramasso, Michèle Rombaut, Denis Pellerin. Modèles de Markov cachés crédibilistes en présence d'informations non distinctes. LFA 2008 - Rencontres Francophones sur la Logique Floue et ses Applications, Oct 2008, Lens, France. hal-00347817

**HAL Id: hal-00347817**

**https://hal.science/hal-00347817v1**

Submitted on 16 Dec 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modèles de Markov Cachés Crédibilistes en présence d'informations non distinctes

# Evidence-theoretic Hidden Markov Models with possibly non distinct information sources

Emmanuel RAMASSO * Michèle ROMBAUT Denis PELLERIN
GIPSA-lab, Images and Signal Department, 46, avenue Félix Viallet, 38031 Grenoble, France
e_ramasso@yahoo.fr, {rombaut, pellerin}@gipsa-lab.inpg.fr

**Résumé :**
Dans le cadre de la reconnaissance de séquences temporelles, nous présentons une version crédibiliste des Modèles de Markov Cachés adaptée au cas où les observations peuvent être non indépendantes. Nous définissons pour cela les procédures *forward*, *backward* et *Viterbi* dans le cadre du Modèle des Croyances Transférables de Ph. Smets. La reformulation est basée sur le Théorème de Bayes Généralisé ainsi que sur les travaux de T. Denœux concernant les opérateurs de combinaisons prudentes. Une application est proposée pour analyser des mouvements humains dans des vidéos d'athlétisme.

**Mots-clés :**
Modèle de Markov Cachés, Reconnaissance de séquences d'états, Modèle des Croyances Transférables, Analyse de mouvements humains, Indexation de vidéos.

**Abstract:**
In the framework of temporal sequence recognition, an evidential version of Hidden Markov Models is proposed in order to combine possibly dependent observations. For that, we redefine the forward, backward and Viterbi procedures in Ph. Smets' Transferable Belief Model. The reformulation is based on the Generalized Bayesian Theorem and on T. Denœux's cautious rules. An application is proposed on human motion analysis in athletics videos.

**Keywords:**
Hidden Markov Models, State sequence recognition, Transferable Belief Model, Human motion analysis, Video Indexing.

## 1 Introduction

Les Modèles de Markov Cachés (HMM) sont des machines à états stochastiques probabilistes permettant de reconnaître des séquences d'états dits *cachés* à partir d'observations bruitées. Les HMM ont été utilisés dans de nombreux domaines tels que la reconnaissance de la parole [7] et l'analyse de vidéos [12].

Les HMM sont basés sur trois mécanismes de base : les algorithmes *forward*, *backward* et *Viterbi* [7]. Dans [8], nous avons proposé une généralisation des HMM aux fonctions de croyance dans le cadre du Modèle des Croyances Transférables (TBM). Cette généralisation permet de bénéficier à la fois de 1) l'efficacité des HMM pour la modélisation et la classification et 2) la généricité du TBM pour la modélisation et la combinaison de connaissances partielles concernant les états. Le TBM permet, par exemple, de modéliser explicitement le doute et le conflit. Le doute peut modéliser des transitions entre états, imprécises par nature. Ce doute est par ailleurs adapté pour la phase d'initialisation de séquences. Le conflit est fortement exploité lors de la classification et représente une source d'information pertinente pour l'analyse de données [8].

Dans cet article, nous proposons une nouvelle version des HMM crédibilistes (CrHMM), plus génériques que ceux proposés dans [8] et où, contrairement à la majorité des approches plus classiques, les sources d'informations peuvent ne pas être distinctes tel que le définit Denœux dans [5].

## 2 HMM probabilistes

Nous supposons que le lecteur possède les connaissances sur les mécanismes de base des HMM décrits dans le tutorial de Rabiner [7].

Les états de la séquence à reconnaître sont dits cachés : on ne dispose que d'un lot de vec-

*E. Ramasso est actuellement en post-doctorat au Laboratoire d'Intelligence Multicapteurs et Apprentissage (LIMA) du CEA / DRT / LIST / DETECS / SSTM.

teurs d'observations $\mathbf{O}_t$ collectées à des instants $t \in \{1..T\}$ ($T$ est la longueur de la séquence) où une observation est un ensemble de $F$ caractéristiques extraites sur le sujet de l'analyse. On note $N$ le nombre d'états et $s_i^t \in \Omega_t$ l'état $i$ à l'instant $t$ faisant partie des $N$ états du cadre du discernement $\Omega_t = \{s_1^t, s_2^t, \dots s_N^t\}$. A chaque instant $t$, la vraisemblance de l'état $i$ conditionnelle à l'observation courante notée $b_{s_i}(\mathbf{O}_t) = P(\mathbf{O}_t|s_i^t)$ est fournie par une technique de modélisation comme par exemple un mélange de Gaussiennes [7].

La transition d'un état $s_i^{t-1}$ à un état $s_j^t$ se fait avec une probabilité $a_{ij} = P^{\Omega_t}(s_j^t|s_i^{t-1})$. La matrice de transition $\mathbb{A} = [a_{ij}], i \in \{1 \dots N\}, j \in \{1 \dots N\}$, de taille $N \times N$ est constituée des probabilités de transition et elle est *stochastique* ($\sum_j a_{ij} = 1$). Les transitions entre états débutent à partir d'une distribution initiale sur les états à l'instant $t = 1$ et notée $\pi_i, i \in \{1 \dots N\}$. Les probabilités sur les états sont mises à jour en fonction des transitions et des observations $b_{s_i}(\mathbf{O}_t)$. Un HMM $\lambda = \{\mathbb{A}, \mathbb{B}, \Pi\}$ est donc caractérisé par trois éléments : la matrice de transition $\mathbb{A} = [a_{ij}], i \in \{1 \dots N\}, j \in \{1 \dots N\}$, les modèles d'observation $\mathbb{B}$ générant les vraisemblances $b_i(\mathbf{O}_t), i \in \{1 \dots N\}, t \in \{1 \dots T\}$ et la distribution initiale $\Pi$ formée des $\pi_i, i \in \{1 \dots N\}$ sur les états à $t = 1$.

La propagation *forward* permet de calculer la variable *forward* $\alpha_t(s_j^t)$ qui représente la vraisemblance d'un état $s_j^t \in \Omega_t$ estimée à partir des observations partielles $\mathbf{O}_{1:t}$ conditionnellement au modèle de HMM $\lambda$ utilisé : $\alpha_t(s_j^t) = P(\mathbf{O}_1, \mathbf{O}_2 \dots \mathbf{O}_t, s_j^t|\lambda) = \sum_{\forall s_i^{t-1} \in \Omega_{t-1}} (\alpha_{t-1}(s_i^{t-1}) \cdot a_{ij}) \cdot b_j(\mathbf{O}_t)$. La propagation *backward* permet quant à elle de calculer la variable *backward* $\beta_t(s_j^t)$ qui représente la vraisemblance des observations partielles $\mathbf{O}_{t+1:T}$ conditionnellement à l'état $s_j^t$ et au modèle de HMM $\lambda$ utilisé : $\beta_t(s_j^t) = P(O_T, O_{T-1} \dots O_{t+1}|s_j^t, \lambda) = \sum_{\forall s_i^{t+1} \in \Omega_{t+1}} (a_{ij} \cdot b_i(\mathbf{O}_{t+1}) \cdot \beta_{t+1}(s_i^{t+1}))$.

La propagation *forward* est utilisée pour la classification de séquences, tandis que deux combinaisons particulières entre les variables *forward* et *backward* sont utilisées dans l'algorithme de Baum-Welch pour l'estimation des paramètres [7]. La première est donnée par : $\gamma_t(s_j^t) = P(s_j^t|\mathbf{O}_{1:T}, \lambda) = \alpha_t(s_j^t) \cdot \beta_t(s_j^t)$ qui est notamment utilisée pour déterminer l'état le plus probable à un instant donné. La seconde est donnée par : $\xi_t(s_i^t, s_j^{t+1}) = P(s_i^t, s_j^{t+1}|\mathbf{O}_{1:T}, \lambda) = \xi_t(s_i^t, s_j^{t+1}) \propto \alpha_t(s_i^t) \cdot a_{ij} \cdot b_j(\mathbf{O}_{t+1}) \cdot \beta_{t+1}(s_j^{t+1})$ et permet d'estimer la probabilité moyenne de transition entre deux états. Notons que l'algorithme dit de *Viterbi* permet de déterminer la séquence d'états la plus vraisemblable et revient à une propagation *forward* où l'opérateur *somme-produit* est remplacé par l'opérateur *max-produit* [7].

## 3 Le TBM : outils de base

La croyance d'un agent sur un ensemble d'hypothèses $\Omega_t$ est représentée par une fonction de croyance [1] qui peut être une distribution de masses (BBA) $m^{\Omega_t}$ de $2^{\Omega_t}$ dans $[0, 1]$, ou d'autres fonctions telles que la plausibilité $pl^{\Omega_t}$ ou la communalité $q^{\Omega_t}$. Par la suite, un singleton est noté $s_i^t \in \Omega_t$ et un sous-ensemble $S_i^t \subseteq \Omega_t$ (nous omettons les accolades). Ces fonctions sont en correspondance biunivoque et les expressions permettant de passer de l'une à l'autre sont données dans [8] (Chap. 1). Par la suite, nous exploiterons la fonction $\omega^{\Omega_t}$ fournissant les poids de la décomposition canonique conjonctive (PDCC) [5] :

$$w^{\Omega}(B) = \prod_{B \subseteq C} q^{\Omega}(C)^{(-1)^{|C|-|B|+1}} \quad \forall B \subseteq \Omega \tag{1}$$

Les poids de la décomposition canonique disjonctive (PDCD) sont liés aux PDCC par $v(B) = \overline{w}(\overline{B})$ [5].

La combinaison de fonctions de croyance peut être effectuée par (Tab. 1) les opérateurs conjonctifs $\textcircled{\cap}$ ou disjonctifs $\textcircled{\cup}$ en utilisant les opérateurs prudents $\textcircled{\wedge}$ et hardis

[1] Le terme croyance est générique et ne représente pas seulement la fonction *bel* que l'on préfère appeler une *crédibilité*.

| sources | toutes fiables | une au moins fiable |
|---|---|---|
| distinctes | $\textcircled{$\cap$}$ | $\textcircled{$\cup$}$ |
| non-distinctes | $\textcircled{$\wedge$}$ | $\textcircled{$\vee$}$ |

**Tableau 1 :** *Choix des règles de fusion [5].*

$\textcircled{$\vee$}$ si les sources de croyance ne sont pas indépendantes [5]. Ces règles sont définies $\forall B \subseteq \Omega$ de la manière suivante :

La **règle de combinaison conjonctive** (CRC) :

$$w^{\Omega}_{1 \textcircled{\cap} 2}(B) = w^{\Omega}_1(B) \cdot w^{\Omega}_2(B) \tag{2}$$

La **règle de combinaison conjonctive prudente** (CCRC) :

$$w^{\Omega}_{1 \textcircled{\wedge} 2}(B) = w^{\Omega}_1(B) \wedge w^{\Omega}_2(B) \tag{3}$$

où $\wedge$ est l'opérateur *minimum*. La définition de cette règle impose que $m^{\Omega}_i(\Omega) > 0$.

Il est possible d'utiliser une infinité d'opérateurs de combinaison de type t-conormes ($\top$) définis dans le cadre de la théorie des possibilités au lieu de l'opérateur $\wedge$ (cf. [5]). Dans l'application proposée, nous avons utilisé la t-conorme de Frank [5] donnée par :

$$x \top_s y = \begin{cases} x \wedge y & \text{si } s = 0 \\ x \cdot y & \text{si } s = 1 \\ \log_s \left( 1 + \dfrac{(s^x - 1) \cdot (s^y - 1)}{s - 1} \right) & \text{sinon} \end{cases} \tag{4}$$

où $s \in [0, 1]$. Lorsque $s = 0$, on retrouve la CCRC et si $s = 1$ on obtient la CRC.

La **règle de combinaison disjonctive prudente** (BDRC) :

$$v^{\Omega}_{1 \textcircled{\vee} 2}(B) = v^{\Omega}_1(B) \wedge v^{\Omega}_2(B) \tag{5}$$

avec $m^{\Omega}_i(\emptyset) > 0$ (BBAs sub-normales [5]).

La prise de décision (après combinaison) est effectuée à partir de la distribution de probabilité pignistique [11] définie $\forall s_k \in \Omega$ par :

$$\mathbf{BetP}\{m^{\Omega}\}(s_k) = C^{-1} \cdot \sum_{B \subseteq \Omega, s_k \in B} \frac{m^{\Omega}(B)}{|B|} \tag{6}$$

$C = (1 - m^{\Omega}(\emptyset))$, $m^{\Omega}(\emptyset)$ est la masse de conflit. Elle peut aussi être effectuée en utilisant la transformée de plausibilités [2] qui possède des caractéristiques différentes de la précédente.

# 4 Algorithmes de base du CrHMM

Comme dans le cas probabiliste, les algorithmes de base du HMM crédibiliste (CrHMM) sont les propagations *forward* et *backward* que nous avons définies dans [8] pour le cas d'observations indépendantes. Nous proposons de reécrire les expressions de ces propagations (en utilisant les PDCC) pour le cas où les observations ne sont pas forcément indépendantes nécessitant les éléments suivants :

- Une connaissance à priori ($\Pi$) à $t = 1$ qui peut être, dans le cadre crédibiliste, la fonction de masse *vide* (toute la masse est placée sur $\Omega$).

- L'ensemble $\mathbb{A}$ des transitions [2] entre sous-ensembles d'états $S^{t-1}_i$ vers les sous-ensembles d'états $S^t_j$ entre deux instants consécutifs.

- L'ensemble $\mathbb{B}$ des paramètres qui permettent de transformer les observations $\mathbf{O}_t$ à l'instant $t$ en fonction de croyance $w^{\Omega_t}_b$.

Nous rappelons que les fonctions de croyance $m$ (masse), $pl$ (plausibilité), $q$ (communalité), $b$ (implicabilité), $bel$ (crédibilité), $w$ (PDCC) et $v$ (PDCD) sont toutes en relation biunivoque.

## 4.1 Algorithme Forward

La propagation *forward* crédibiliste fait appel aux éléments suivants :

- la BBA de la variable *forward* à l'instant précédent $m^{\Omega_{t-1}}_{\alpha}$ à $t - 1$,

- les PDCC $w^{\Omega_t}_a[S^{t-1}_i]$ définis sur $2^{\Omega_t}$ conditionnellement aux sous-ensembles d'états précédents ($S^{t-1}_i \subseteq \Omega_{t-1}$), et définissant les liens entre sous-ensembles d'états à $t - 1$ et $t$,

- $w^{\Omega_t}_b$ qui décrit l'information provenant des ob-

[2] La matrice de transition est généralement de taille $N \times 2^N$ au lieu de $2^N \times 2^N$ car, comme précisé par Ph. Smets [9], on dispose généralement de croyances conditionnelles aux singletons $s_i$ à partir desquelles les croyances conditionnelles aux sous-ensembles $S$ sont obtenues par combinaison disjonctive.

servations à $t$ (équivalent à $b(O_t)$).

La propagation *forward* crédibiliste suit le schéma classique « prédiction - mise à jour » :
**Étape 1 :** La prédiction $\hat{w}_\alpha^{\Omega_t}(S_j^t)$ consiste à combiner la BBA sur les propositions $S_j^t$ avec les PDCC en utilisant l'équation 7 :

$$\hat{w}_\alpha^{\Omega_t}(S_j^t) = \sum_{S_i^{t-1} \subseteq \Omega_{t-1}} m_\alpha^{\Omega_{t-1}}(S_i^{t-1}) \cdot w_a^{\Omega_t}[S_i^{t-1}](S_j^t) \quad (7)$$

Cette relation estime la somme des contributions des éléments focaux à $t-1$ pondérée par les transitions $w_a^{\Omega_t}[S_i^{t-1}]$ de $S_i^{t-1}$ à $S_j^t$.

**Étape 2 :** Mise à jour de la prédiction par les observations à $t$. Quand celles-ci ne sont pas indépendantes, on utilise la CCRC :

$$w_\alpha^{\Omega_t}(S_j^t) = \hat{w}_\alpha^{\Omega_t}(S_j^t) \wedge w_b^{\Omega_t}(S_j^t) \quad (8)$$

Si les observations sont indépendantes, l'opérateur $\wedge$ est remplacé par un produit.

**Remarque 1 :** La propagation *forward* peut débuter avec $w_\alpha^{\Omega_1}(S_i^1) = 1, \forall S_i^1 \subseteq \Omega_1$ (ignorance totale à l'instant initial).

**Remarque 2 :** La combinaison entre observations et prédiction peut engendrer un conflit :

$$m_\alpha^{\Omega_t}(\emptyset) = 1 - pl_\alpha^{\Omega_t}(\Omega_t) \quad (9)$$

Le conflit à $t$ doit être traité afin qu'il ne soit pas propagé au temps $t+1$ car il est absorbant par l'opérateur conjonctif. Pour cela il est possible d'utiliser l'une des règles de redistribution proposées dans la littérature [10].

**Remarque 3 :** La propagation *forward* crédibliste généralise le cas probabiliste si les observations et les transitions sont Bayésiennes (i.e. de la masse est présente seulement sur les singletons) et si la combinaison de Dempster [3] est utilisée [4].

### 4.2 Algorithme Backward

L'algorithme *backward* est très similaire au *forward* et nécessite :

- la BBA $m_\beta^{\Omega_{t+1}}$ de la variable *backward* définie sur les sous-ensembles d'états à $t+1$,

- l'ensemble des plausiblités conditionnelles $pl_a^{\Omega_{t+1}}[S_i^t]$, qui lient les états de $t$ à $t+1$,

- la BBA $m_b^{\Omega_{t+1}}$ obtenue à partir des observations à $t+1$.

On peut alors estimer la fonction de croyance a posteriori sur $\Omega_t$ de la façon suivante :

**Étape 1 :** Calcul de la BBA a priori sur $\Omega_{t+1}$ par combinaison conjonctive de $m_\beta^{\Omega_{t+1}}$ et $m_b^{\Omega_{t+1}}$. Si les observations sont indépendantes, on utilise la combinaison CRC (et on obtient $m_{\beta \textcircled{\cap} b}^{\Omega_{t+1}}$) sinon on utilise la CCRC (menant à $m_{\beta \textcircled{\wedge} b}^{\Omega_{t+1}}$).

**Étape 2 :** Projection de $m_{\beta \textcircled{\cap} b}^{\Omega_{t+1}}$ ou $m_{\beta \textcircled{\wedge} b}^{\Omega_{t+1}}$ sur $\Omega_t$. Ceci nécessite de calculer la communalité a posteriori $q_a^{\Omega_t}[S_j^{t+1}]$ à partir des plausibilités $pl_a^{\Omega_{t+1}}[S_i^t]$ en utilisant un résultat du Théorème de Bayes Généralisé (GBT) [9] :

$$q_a^{\Omega_t}[S_j^{t+1}](S_i^t) = \prod_{s_i^t \in S_i^t} pl_a^{\Omega_{t+1}}[s_i^t](S_j^{t+1}) \quad (10)$$

La communalité obtenue à l'équation 10 est ensuite transformée en PDCC $w_a^{\Omega_t}[S_j^{t+1}]$ en appliquant l'équation 1. Puis, ces PDCC conditionnels sont utilisés pour pondérer la projection de la masse obtenue à l'étape 1 vers les sous-ensembles d'états à $t$. Quand les observations ne sont pas indépendantes, on obtient :

$$w_\beta^{\Omega_t}(S_i^t) = \sum_{S_j^{t+1} \subseteq \Omega_{t+1}} m_{\beta \textcircled{\wedge} b}^{\Omega_{t+1}}(S_j^{t+1}) \cdot w_a^{\Omega_t}[S_j^{t+1}](S_i^t) \quad (11)$$

La distribution $w_\beta^{\Omega_T}$ est *a priori* inconnue et donc initialisée à une croyance vide : $w_\beta^{\Omega_T}(S_i^T) = 1, \forall S_i^T \subseteq \Omega_T$.

**Remarque 4 :** La propagation *backward* crédibiliste généralise le cas probabiliste si les BBA sont Bayésiennes et si la CRC est utilisée avec normalisation. Ce résultat est assuré grâce au Théorème de Bayes Généralisé [9].

[3] Cette combinaison est une CRC suivie d'une normalisation.

# 5 Fonctionnalités des CrHMM

## 5.1 Filtrage de fonctions de croyance

Le filtrage en ligne des fonctions de croyance sur les états peut être réalisé en utilisant la propagation *forward*. En effet cette dernière permet d'assurer une cohérence temporelle sur les croyances grâce aux transitions. De plus, et de la même manière que dans le cadre probabiliste (et crédibiliste dans le cas indépendant [8]), il est possible de combiner les propagations *forward* et *backward* pour lisser (hors ligne) des fonctions de croyance non indépendantes. Pour cela, on définit la fonction de croyance $f_\gamma^{\Omega_t}$, équivalent crédibiliste de la fonction $\gamma$ définie dans les HMM probabilistes [7]. Le lissage hors ligne est obtenue en combinant de façon conjonctive les variables *forward* et *backward*. Si les observations ne sont pas indépendantes, on obtient :

$$w_\gamma^{\Omega_t}(S_j^t) = w_\alpha^{\Omega_t}(S_j^t) \wedge w_\beta^{\Omega_t}(S_j^t) \qquad (12)$$

## 5.2 Détermination de l'état du système

Pour un HMM donnée, l'état $s_t^*$ le plus probable à $t$ est trouvé en maximisant la probabilité pignistique (Eq. 6) $\mathbf{BetP}\{m_\gamma^{\Omega_t}\}$ calculée à partir de la BBA $m_\gamma^{\Omega_t}$ ($\forall t \in \{1 \dots T\}$) :

$$s_t^* = \underset{s_i \in \Omega_t}{\operatorname{argmax}} \ \mathbf{BetP}\{m_\gamma^{\Omega_t}\}(s_i) \qquad (13)$$

Comme dans le cas probabiliste [7], cette technique permet d'obtenir la séquence d'états dont la probabilité est maximum à un instant donné mais elle ne prend pas en compte la séquence des états dans la maximisation à la différence de l'algorithme de Viterbi (§ 5.4).

## 5.3 Classification d'observations

Le problème de classification dans le cadre des CrHMM est le suivant : pour une séquence donnée d'observations $\mathbf{O}_{1:T}$ et pour un ensemble de $N$ CrHMM candidats $\lambda_1, \dots, \lambda_n, \dots, \lambda_N$, il s'agit de choisir le CrHMM qui correspond au mieux aux observations.

Pour cela, reconsidérons l'étape de mise à jour de la propagation *forward* (Eq. 8). La combinaison conjonctive qu'elle implique entre prédiction et observations peut générer un conflit interprété de la façon suivante : *plus le conflit est faible tout au long de la séquence d'observations, plus le modèle $\lambda$ du CrHMM correspond à ces observations*.

Pour un modèle $\lambda$, la valeur de la variable *forward* crédibiliste pour une proposition $B \subset \Omega_t$ dans l'espace des BBA $m$ est donnée par :

$$m_{\alpha,*}^{\Omega_t}[\lambda](B) = C^{-1} \cdot m_\alpha^{\Omega_t}[\lambda](B) \qquad (14)$$

où $C = pl_\alpha^{\Omega_t}[\lambda](\Omega_t) = 1 - m_\alpha^{\Omega_t}[\lambda](\emptyset) = \sum_{D \subseteq \Omega_t, D \neq \emptyset} m_\alpha^{\Omega_t}[\lambda](D)$. Cette normalisation est équivalente à celle utilisée dans le cadre probabiliste (voir équation 100 dans [7]). Cependant, dans le cas crédibiliste, la normalisation peut être interprétée de façon particulièrement intéressante puisque le coefficient de normalisation correspond au conflit entre la prédiction (et donc validité du modèle) et les observations [4].

Ainsi, pour classer une séquence d'observations $\mathbf{O}_{1:T}$, on commence par calculer le critère qui intègre le conflit au cours du temps (on peut utiliser le logarithme lors de l'implémentation) :

$$\mathcal{L}_c(\lambda) = \frac{1}{T} \sum_{t=1}^{T} \log(m_\alpha^{\Omega_t}[\lambda](\emptyset)) \qquad (15)$$

ou encore $1/T \cdot \sum_{t=1}^{T} \log(1 - pl_\alpha^{\Omega_t}[\lambda](\Omega_t))$ d'après l'équation 9. Il reste alors à choisir le CrHMM $\lambda_*$ qui *miminise* ce critère pour la séquence d'observations considérée :

$$\lambda^* = \underset{n}{\operatorname{argmin}} \ \mathcal{L}_c(\lambda_n) \qquad (16)$$

Ce critère généralise le critère probabiliste, sauf que, dans le cas probabiliste, on cherche à maximiser la vraisemblance qui est ici équivalente à $pl_\alpha^{\Omega_t}[\lambda](\Omega_t)$.

[4] On voit par ailleurs que le conflit est lié à l'opposé de la plausibilité de l'ensemble global qui est équivalente, dans le cas probabiliste (i.e. si les croyances sont bayésiennes), à la vraisemblance des observations conditionnellement aux états.

En conclusion, l'algorithme *forward* crédibiliste pour la classification de séquence (ou reconnaissance de formes) consiste en 5 étapes : 1) la première BBA est initialisée à la BBA vide, 2) on applique l'équation 8 à $t$, 3) on mémorise le conflit à $t$, 4) on redistribue le conflit et 5) on calcule la valeur du critère 15. Ce processus est itéré pour $t = 1$ à $T$ pour chaque CrHMM. On détermine ensuite le meilleur modèle par l'équation 16.

### 5.4 Meilleure séquence d'états (Viterbi)

Pour un HMM donné, l'objectif est de déterminer la meilleure séquence d'états $s^1, s^2, \dots s^T$, $s^t \in \Omega_t$ étant donnée la séquence d'observations $\mathbf{O}_{1:T}$. Dans le cas probabiliste, l'algorithme de Viterbi a été proposé pour cette tâche. Il permet de sélectionner la séquence ayant le maximum de vraisemblance sans calculer l'ensemble des $N^T$ séquences possibles. Cet algorithme consiste à calculer [7] :

$$\begin{aligned}\delta_t(s_j^t) &= \max_{s^1, s^2 \dots s^{t-1}} P(s^1, s^2 \dots s_j^t, \mathbf{O}_{1:t}|\lambda) \\ &= \max_{\forall s_i^{t-1} \in \Omega_{t-1}} \left[\delta_{t-1}(s_i^{t-1}) \cdot a_{ij}\right] \cdot b_j(\mathbf{O}_t) \\ \psi_t(s_j^t) &= \operatorname*{argmax}_{\forall s_i^{t-1} \in \Omega_{t-1}} \left[\delta_{t-1}(s_i^{t-1}) \cdot a_{ij}\right]\end{aligned}$$

La variable $\psi_t$ permet de mémoriser les prédécesseurs de chaque état $s_j^t$ à chaque itération. Lorsque l'ensemble de la séquence est traitée, on choisit l'état à $T$ qui maximise $\delta_T(s_j^T)$ puis on rétropropage cette décision jusqu'à $t = 1$ par $s_t^* = \psi_{t+1}(s_{t+1}^*)$.

Dans le cadre crédibiliste, la métrique de Viterbi propagée est une fonction de croyance notée $w_\delta^{\Omega_t}$ définie à chaque instant $t$ sur $\Omega_t$. Notons que, dans sa version probabiliste, l'algorithme propage les scores de décisions (la vraisemblance maximum) à chaque instant. Dans le cas crédibiliste, la décision concernant le choix des meilleurs prédécesseurs devra elle-aussi être effectuée sur les états singletons et à l'aide de la probabilité pignistique (Eq. 6). Ces décisions étant sur les singletons, la propagation de ces dernières à l'instant suivant ferait perdre l'intérêt d'utiliser le cadre des fonctions de croyance car la combinaison d'une fonction de croyance Bayésienne [5] avec une fonction croyance quelconque génére une fonction de croyance Bayésienne. Dans [8] (Chap. 5), nous avons proposé plusieurs alternatives permettant de se maintenir dans le cadre des fonctions de croyance. L'alternative retenue qui a montré des résultats théoriques et applicatifs intéressants est étendue ci-après pour le cas où les observations ne sont plus indépendantes.

La croyance $w_\delta^{\Omega_t}$ est construite en quatre étapes :

**Étape 1 :** Prédiction de la fonction de croyance $\hat{w}_{\delta,i}^{\Omega_t}$ à $t$ à partir de la métrique à l'instant précédent $m_\delta^{\Omega_{t-1}}$ ($\forall s_i^{t-1} \in \Omega_{t-1}, \forall B \subseteq \Omega_t$) :

$$\hat{w}_{\delta,i}^{\Omega_t}(B) = \sum_{S \subseteq \Omega_{t-1}} w_a[S \cap s_i](B) \cdot m_\delta^{\Omega_{t-1}}(S) \quad (18)$$

**Étape 2 :** Mise à jour de la prédiction $\hat{w}_{\delta,i}^{\Omega_t}$ avec les observations $w_b^{\Omega_t}$. Si les observations ne sont pas indépendantes, on a :

$$w_{\delta,i}^{\Omega_t}(B) = \hat{w}_{\delta,i}^{\Omega_t}(B) \wedge w_b^{\Omega_t}(B) \quad (19)$$

et si les observations sont indépendantes, nous utilisons un produit au lieu de $\wedge$.

**Étape 3 :** Construction de la probabilité pignistique à partir de la croyance $w_{\delta,i}^{\Omega_t}$ :

$$\mathcal{P}_t[s_i^{t-1}](s_j^t) = \mathbf{BetP}^{\Omega_t}\{w_{\delta,i}^{\Omega_t}\}(s_j^t) \quad (20)$$

La prédiction (Eq. 18) fait intervenir un conditionnement sur $s_i^{t-1} \times \Omega_t$ ce qui est équivalent à une combinaison conjonctive avec une BBA catégorique [6]. Le conflit engendré reflète la pertinence de l'hypothèse « *le prédécesseur de* $s_j^t$ *est* $s_i^{t-1}$ » et la valeur de conflit est mémorisée :

$$\Gamma_t(s_i^{t-1}) = 1 - m_{\delta,i}^{\Omega_t}(\emptyset) \quad (21)$$

*Moins le conflit est grand, plus l'hypothèse « le prédécesseur de* $s_j^t$ *est* $s_i^{t-1}$ *» est vraie* et cette information mémorisée sera ensuite utiliser pour trouver le meilleur prédécesseur.

[5] La fonction de croyance obtenue à partir des décisions.

[6] Une BBA catégorique possède un seul élément non nul et donc de masse unitaire.

**Étape 4 :** Quand les étapes 1 à 3 sont effectuées *pour tous les états prédécesseurs* $s_i^{t-1} \in \Omega_{t-1}$, le prédécesseur le plus probable de $s_j^t \in \Omega_t$ est choisi en maximisant la probabilité pignistique calculée à l'étape 3 :

$$\psi_t^{'}(s_j^t) = \underset{s_i^{t-1} \in \Omega_{t-1}}{\text{argmax}} \left[\Gamma_t(s_i^{t-1}) \cdot \mathcal{P}_t[s_i^{t-1}](s_j^t)\right] \quad (22a)$$

$$\delta_t^{'}(s_j^t) = \Gamma_t(\psi_t^{'}(s_j^t)) \cdot \mathcal{P}_t[\psi_t^{'}(s_j^t)](s_j^t) \quad (22b)$$

où $\psi_t^{'}(s_j^t)$ et $\delta_t^{'}(s_j^t)$ correspondent respectivement au meilleur prédécesseur et la valeur maximale de la transition de $s_j^t$ à $s_j^t$.

**Étape 5 :** La propagation de la maximisation est effectuée par une *propagation forward conditionnée* de $w_\delta^{\Omega_{t-1}}$. Le conditionnement est réalisé sur l'union des prédécesseurs définie par $\mathcal{A}_{t-1} = \bigcup_{s_j^t \in \Omega_t} \psi^{'}(s_j^t)$. Quand les observations ne sont pas indépendantes, la métrique de Viterbi à $t$ est définie $\forall S_j^t \subseteq \Omega_t$ par :

$$w_\delta^{\Omega_t}(S_j^t) = w_b^{\Omega_t}(S_j^t) \wedge \sum_{S \subseteq \Omega_{t-1}} m_\delta^{\Omega_{t-1}}(S) \cdot w_a^{\Omega_t}[S \cap \mathcal{A}_{t-1}](S_j^t) \quad (23)$$

et si les observations sont indépendantes, l'opérateur produit remplace $\wedge$. La stratégie proposée est un peu différente de celle utilisée en probabilité car elle ne propage pas directement la décision mais uniquement les composantes intéressantes (sur les prédécesseurs) de $w_\delta^{\Omega_{t-1}}$ pondérées par les croyances conditionnelles. Cette stratégie, bien que sous optimale dans le sens où l'on ne maximise pas de critère explicite, permet de tirer parti de la modélisation par fonctions de croyance.

## 6 Illustration

Nous proposons d'illustrer l'utilisation de l'algorithme de Viterbi crédibiliste sur une application de reconnaissance de mouvements humains dans des vidéos. Nous utilisons des données décrites dans [8] concernant la reconnaissance de 15 sauts en hauteur. Un saut en hauteur est modélisée par un CrHMM composés de 4 états avec $\Omega_t = \{s_1^t = \text{course}, s_2^t = \text{saut}, s_3^t = \text{chute}, s_4^t = \text{relève}\}$. A chaque image, des caractéristiques (observations : $\mathbf{O}_t \in \Re^F$) sont extraites concernant chaque état. L'apprentissage des modèles d'observations est effectué pour chaque état à l'aide de l'algorithme ECM [7] proposé par M. Masson et T. Denœux [6]. A chaque état $s_j$, on associe un espace $\Omega_{ecm}^j$ formé de cinq éléments correspondant à des nuages de points d'apprentissage dans l'espace des observations. Lorsque ECM a convergé, il permet d'obtenir une fonction de croyance $m^{\Omega_{ecm}^j}$ qui modélise une partition évidentielle sur l'ensemble d'apprentissage [8]

Quand une nouvelle observation $\mathbf{O}_t$ est disponible, il s'agit ensuite de déterminer en quoi elle correspond à l'état $s_j$. Dans le cas où elle correspond peu, donc en dehors des nuages de points, on obtient une masse sur l'ensemble vide (monde ouvert) : $m^{\Omega_{ecm}^j}(\emptyset) = 1 - \sum_B m^{\Omega_{ecm}^j}(B) = 1 - pl^{\Omega_{ecm}^j}(\Omega_{ecm}^j)$ (Eq. 9). Cette masse est interprétée de la manière suivante : *plus le conflit obtenu est important, moins l'observation est cohérente avec la partition*. La quantité $pl^{\Omega_{ecm}^j}(\Omega_{ecm}^j)$ peut alors être vue comme la plausibilité de l'observation $\mathbf{O}_t$ sachant l'état $j$, i.e. $pl^{\Re^F}[s_j^t](\mathbf{O}_t)$, qui peut être traitée par le GBT pour obtenir la BBA a posteriori comme dans le cas du modèle de vraisemblance [3] :

$$m^{\Omega_t}[\mathbf{O}_t](S^t) = \prod_{s_j^t \in S^t} pl^{\Re^F}[s_j^t](\mathbf{O}_t) \times \prod_{s_j^t \notin S^t} (1 - pl^{\Re^F}[s_j^t](\mathbf{O}_t)) \quad (24)$$

Ces masses représentent les entrées de l'algorithme de Viterbi. Figure 1, les 4 premières courbes correspondant aux entrées ne permettent pas de différencier les états par le HMM, mais l'information est suffisante pour le CrHMM. Les matrices de transition

[7] Code disponible sur la page web de T. Denœux.

[8] Paramétrage de l'ECM : $\alpha = 1$, $\beta = 2$, $\delta = 150$, $\epsilon = 1e-3$ et $M = 5$ nombre de nuages de points (*clusters*) choisi identique pour tous les états.

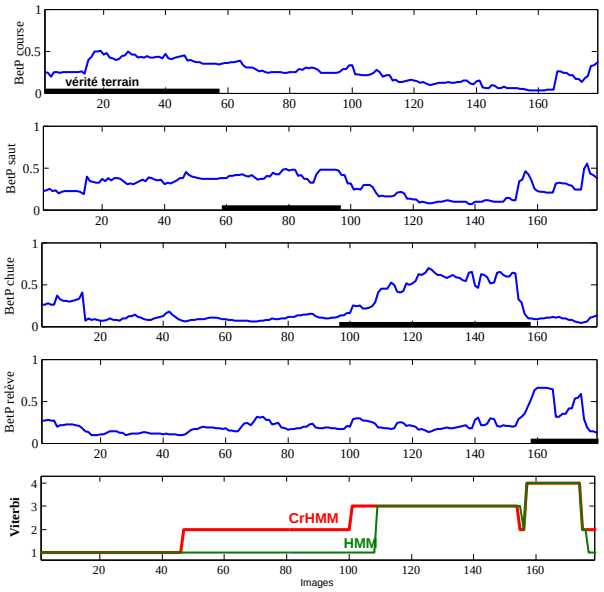


**Figure 1 :** *Résultat de l'algorithme de Viterbi crédibiliste vs. probabiliste (tracé 5) et évolution des probabilités pignistiques sur les états (par Eq. 24 suivie de Eq. 6).*

| | CrHMM-1 | CrHMM-2 | CrHMM-3 |
|---|---|---|---|
| Co | +2 | +4 | +6 |
| Sa | +14 | +16 | +19 |
| Ch | +1 | +6 | +8 |
| Rel | +1 | +2 | +6 |

**Tableau 2 :** *Comparaison des performances (en %) de la détection des états (Co : course, Sa : saut, Ch : chute et Rel : se relever dans des sauts en hauteur) par l'algorithme de Viterbi crédibiliste vs. probabiliste (un signe « + » signifie que la performance du CrHMM est meilleure que celle du HMM). Réglages ($s$ est le paramètre de la t-conorme de Frank) : CrHMM-1 avec $s = 0$, CrHMM-2 avec $s = 1$, CrHMM-3 avec $s = 0.35$.*

des CrHMM sont estimées à l'aide d'une méthode proposée dans [8] qui est similaire au mécanisme probabiliste mais étendu aux fonctions de croyance.

Le tableau 2 présente la différence moyenne de performance [9] entre un HMM [10] et trois versions de CrHMM pour la détection des 4 états. Les CrHMM diffèrent par le réglage du paramètre $s$ (t-conorme de Frank) et montrent ici l'intérêt des opérateurs prudents.

[9] La performance est donnée par la mesure $F_1$ couramment utilisée dans en multimédia et qui combine la rappel et la précision.

[10] Le HMM probabiliste équivalent au CrHMM est obtenu par application de la transformée pignistique (Eq. 6) sur 1) la BBA obtenue à l'équation 24 (une illustration est donnée sur les 4 premiers tracés de la figure 1), 2) la matrice de transition et 3) l'*a priori*.

# 7 Conclusion

Les algorithmes *forward-backward* et *Viterbi* des Modèles de Markov Cachés Crédibilistes (HMM) permettent de combiner des observations éventuellement non indépendantes. Cette nouvelle formulation permet potentiellement de bénéficier des nombreux opérateurs proposés dans la théorie des possibilités.

Divers travaux futurs sont à mener, par exemple, il serait intéressant d'étudier l'impact du choix des opérateurs de combinaison dans le cas non indépendant voire de proposer une méthode pour un choix automatique ou guidé. Enfin, à la suite de l'apprentissage des modèles d'observation à l'aide d'ECM que nous avons proposé dans cet article, nos travaux actuels s'orientent vers l'apprentissage des CrHMM en intégrant les travaux de E. Côme et al. [1].

# Références


[1] E. Côme, L. Oukhellou, T. Denoeux, and P. Aknin. Mixture model estimation with soft labels. In *SMPS*, Toulouse, 2008.

[2] B.R. Cobb and P.P. Shenoy. On the plausibility transformation method for translating belief function models to probability models. *Int. Jour. of Approximate Reasoning*, 41(3) :314–330, 2006.

[3] F. Delmotte and Ph. Smets. Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model. *IEEE Trans. on Systems, Man and Cybernetics*, 34(4), July 2004.

[4] A.P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society*, 30 :205–247, 1968.

[5] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. *Artificial Intelligence*, 2007. To appear.

[6] M.-H. Masson and T. Denoeux. Ecm : An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41 :1384–1397, 2008.

[7] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77 :257–285, 1989.

[8] E. Ramasso. *Reconnaissance de séquences d'états par le Modèle des Croyances Transférables et application à l'analyse de vidéos d'athlétisme*. PhD thesis, Université Joseph Fourier - Grenoble, 5 Décembre 2007.

[9] Ph. Smets. Beliefs functions : The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *Int. Jour. of Approximate Reasoning*, 9 :1–35, 1993.

[10] Ph. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8(4) :387–412, October 2005.

[11] Ph. Smets. Decision making in the TBM : The necessity of the pignistic transformation. *Int. Jour. of Approximate Reasoning*, 38 :133–147, 2005.

[12] A. Wilson and A. Bobick. Hidden Markov models for modeling and recognizing gesture under variation. *Int. Jour. of Pattern Recognition and Artificial Intelligence*, 15(1) :123–160, 2001.